

基于动态样本相似度的多任务学习式跟踪方法

史再峰^{1*}, 孙诚^{1**}, 曹清洁², 王哲¹, 樊强强¹

¹天津大学微电子学院, 天津 300072;

²天津师范大学数学科学学院, 天津 300072

摘要 针对在线更新式跟踪方法易受嘈杂样本干扰进而产生漂移现象的问题, 在跟踪流程中加入丢失检测步骤, 并结合多任务学习训练方式, 提出一种适合长期跟踪的方法。该方法在跟踪过程中长期收集目标外观来构建动态样本库, 根据样本相似度检测目标的丢失, 并使用动态阈值适应不同目标, 来减少跟踪器对嘈杂样本的学习。为使跟踪器建立完整的目标外观模型, 本文联合训练短期记忆子任务和长期记忆子任务。在丢失目标后的重检测过程中, 根据区域轮廓特征和目标尺度信息进行区域提议, 以提高目标重检测的质量。该方法在目标跟踪数据集 OTB-2015 和 VOT-2016 上进行了评估, 其精确度为 90.8%, 成功率为 68.1%。实验结果表明, 在遮挡等复杂场景下该方法可以有效地跟踪目标。

关键词 机器视觉; 在线更新; 目标跟踪; 多任务学习

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.1615002

Multi-Task Learning Tracking Method Based on the Similarity of Dynamic Samples

Shi Zaifeng^{1*}, Sun Cheng^{1**}, Cao Qingjie², Wang Zhe¹, Fan Qiangqiang¹

¹School of Microelectronics, Tianjin University, Tianjin 300072, China;

²School of Mathematical Sciences, Tianjin Normal University, Tianjin 300072, China

Abstract To solve the problem of noisy samples easily interfering with the online updating tracking method and resulting in a drift phenomenon, a method suitable for long-term tracking is proposed, and the proposed method is combined with a multi-task learning training mode and a loss detection step is added into the tracking process. The proposed method constantly collects the appearance of the target during tracking to construct a dynamic sample set, which detects the loss of target according to sample similarity to reduce the tracker's learning of noisy samples; further, the dynamic threshold is used to adapt to different targets. To make the tracker build a complete model of the target appearance, short- and long-term memory subtasks are jointly trained. During redetection, after the target is lost, regions are proposed based on regional outline features and scale information about the target to improve the quality of target redetection. The proposed method is evaluated on the object tracking datasets OTB-2015 and VOT-2016, and the tracker has an accuracy of 90.8% and a success rate of 68.1%. Experimental results show that the proposed method can effectively track a target in complex scenes, such as occlusion.

Key words machine vision; online updating; object tracking; multi-task learning

OCIS codes 150.0155; 150.1135; 100.4999

收稿日期: 2020-10-22; 修回日期: 2020-11-13; 录用日期: 2020-12-08

基金项目: 国家自然科学基金(61674115, 62071326)

通信作者: * shizaifeng@tju.edu.cn; ** suncheng@tju.edu.cn

1 引言

目标跟踪是计算机视觉领域中的重要问题,被广泛应用于人机交互、视频压缩等领域^[1]。当前,跟踪器受到的挑战主要来自复杂的跟踪场景和目标运动,如照明变化、背景杂乱、运动模糊等^[2]。在这些挑战因素的干扰下进行稳定的跟踪是目标跟踪领域的重要研究方向。

早期的目标跟踪器着重于短时间内的本地跟踪任务。文献[3-5]提出了具有代表性的本地跟踪器,随后研究者从运动模型^[6-7]、特征提取^[8-10]、模型更新^[11]等方面进行改进,在短期跟踪任务中取得了良好的表现。然而这些方法在长期的跟踪场景中缺乏稳定的性能,当前长期跟踪稳定性的问题成为目标跟踪领域的研究热点。Kalal 等^[12]在判别类跟踪器中加入检测模块来快速搜索丢失的目标,提出了跟踪学习检测(TLD)方法。文献[13]提出一种浏览模块来快速、粗略地搜索目标存在的区域。文献[14]使用支持向量机(SVM)进行分类来实现重检测,并通过被动攻击算法更新分类器。文献[15]根据相关滤波器的响应检测目标是否被遮挡或丢失,以决定是否进行全图搜索。上述方法通过改进重检测、丢失检测等步骤提高了方法的长期跟踪能力,然而跟踪器难以长期准确判断跟踪结果的真伪,在线更新的跟踪器会不断受到嘈杂样本的污染,使

其难以在长期维度上适应目标外观的变化,进而导致跟踪器漂移甚至永久丢失目标。

针对以上问题,本文提出了一种基于动态样本相似度和多任务学习的跟踪方法。由于本地跟踪器输出的置信度不完全可靠,本文根据样本间的余弦相似度和反映目标变化剧烈程度的动态阈值检测异常的跟踪结果,以减少嘈杂样本对跟踪器的污染。为使跟踪器适应长期跟踪过程中目标外观的各种变化,本文通过多任务学习平衡跟踪器对目标短期外观和长期外观的记忆。在目标丢失后的全图搜索过程,根据轮廓特征和尺度信息进行区域提议,来提供高质量的提议区域。在跟踪基准上的实验表明,本文方法在复杂因素的干扰下可以长期、稳定地跟踪目标。

2 本文跟踪方法

2.1 跟踪方法总体流程

在跟踪器漂移或目标消失时会收集到偏离目标区域的嘈杂样本^[16],给在线更新引入错误。因此要检测不可信的跟踪结果,减少对嘈杂样本的收集。同时,跟踪器不断从数据非平稳分布的样本中学习目标外观,这往往会导致其对早期目标特征的遗忘^[17]。为了建立完整的目标外观模型,需要在更新时恢复跟踪器对早期目标特征的记忆。

跟踪方法的总体流程如图 1 所示。本文使用多

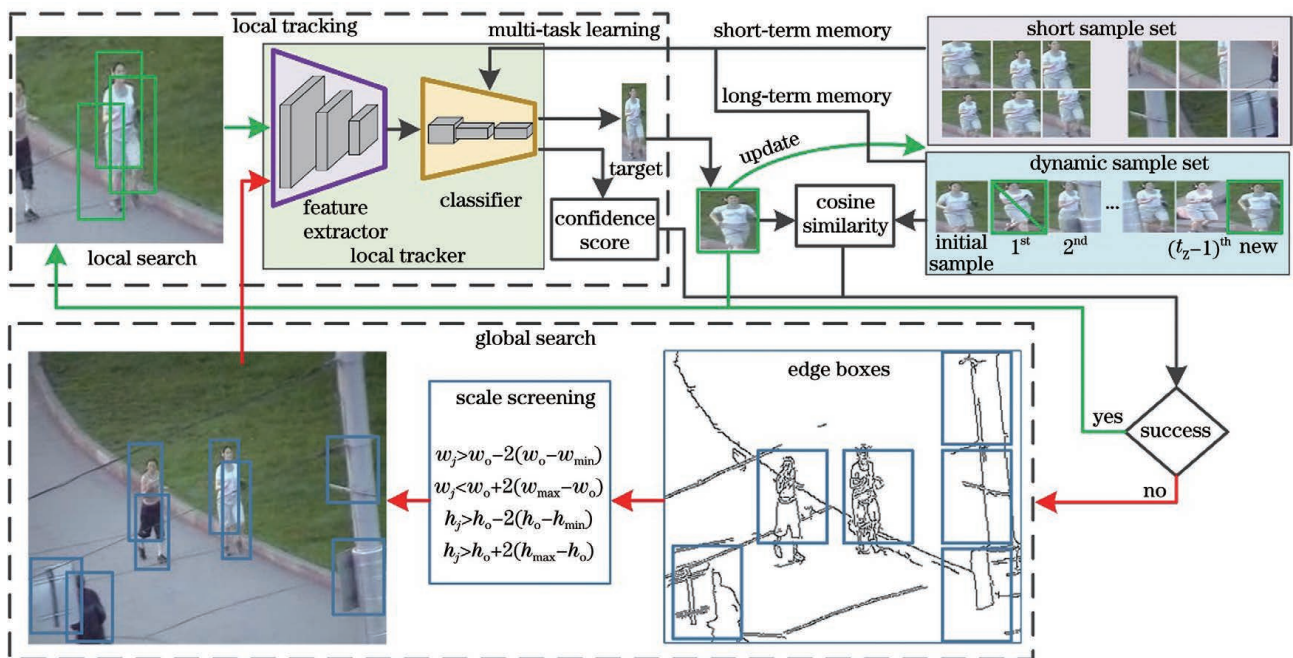


图 1 跟踪方法流程图

Fig. 1 Flow chart of the tracking method

域卷积神经网络 (MDNet) 作为本地跟踪器。在进行每一帧跟踪时, 首先进行候选区域提议, 并由本地跟踪器评估候选样本来得到当前帧的目标位置和置信度得分。然后, 根据置信度得分和目标样本与动态样本的相似度判断是否丢失目标。如果跟踪成功, 则将跟踪结果的图像特征放入动态样本库, 并在下一帧进行局部的区域提议, 否则考虑进行全图范围的候选区域提议。最后, 在跟踪过程中以多任务学习方式定期更新网络模型。

2.2 动态样本相似度

由于跟踪器自身输出的置信度不完全准确, 本文收集目标外观构建动态样本库, 以样本间的余弦

相似度作为检测目标丢失的依据。首先需要获得具有可靠性和多样性的目标外观, 本文将成功的跟踪结果样本添加到动态样本库中。跟踪成功需满足以下条件,

$$\begin{cases} c_{\text{on}} > 0 \\ s_{\text{cos}} > H_{\text{cos}} \end{cases}, \quad (1)$$

式中: c_{on} 是本地跟踪器 MDNet 对跟踪结果的置信度分数, 一般以 0 为可信与不可信的分界线; s_{cos} 是目标样本与动态样本库中各样本的余弦相似度的最大值; H_{cos} 是相似度阈值。

为了比较样本之间高层次的区别, 使用特征提取网络提取出的语义特征计算样本之间的最大余弦相似度, 表示为

$$s_{\text{cos}} = \max \left[\frac{\mathbf{f}^T(\mathbf{Z}_{\text{cs}}) \cdot \mathbf{f}(\mathbf{B}_i)}{\|\mathbf{f}(\mathbf{Z}_{\text{cs}})\|_2 \cdot \|\mathbf{f}(\mathbf{B}_i)\|_2}, \frac{\mathbf{f}^T(\mathbf{Z}_{\text{cs}}) \cdot \mathbf{f}(\mathbf{B}_o)}{\|\mathbf{f}(\mathbf{Z}_{\text{cs}})\|_2 \cdot \|\mathbf{f}(\mathbf{B}_o)\|_2} \right], \quad (2)$$

式中: \mathbf{f} 为特征值提取结果的向量; \mathbf{Z}_{cs} 为当前帧跟踪结果区域的样本; \mathbf{B}_i 为动态样本库中的第 i 个动态样本; \mathbf{B}_o 为第一帧的初始样本。

传统的跟踪方法采用固定阈值判断异常的发生^[13]。由于不同跟踪目标外观变化的剧烈程度不

同, 固定的阈值难以同时适合于不同的跟踪场景, 需要根据跟踪目标调整参数。本文提出动态阈值来解决这一问题, 根据跟踪目标历史外观的相似度来确定阈值 H_{cos} , 表示为

$$H_{\text{cos}} = \frac{1}{t_{\text{ps}}(t_{\text{ps}} + 1)} \left[\sum_{i=1}^{t_{\text{ps}}} \sum_{k=1(k \neq i)}^{t_{\text{ps}}} \frac{\mathbf{f}^T(\mathbf{B}_k) \cdot \mathbf{f}(\mathbf{B}_i)}{\|\mathbf{f}(\mathbf{B}_k)\|_2 \cdot \|\mathbf{f}(\mathbf{B}_i)\|_2} + 2 \times \sum_{i=1}^{t_{\text{ps}}} \left(\frac{\mathbf{f}^T(\mathbf{B}_o) \cdot \mathbf{f}(\mathbf{B}_i)}{\|\mathbf{f}(\mathbf{B}_o)\|_2 \cdot \|\mathbf{f}(\mathbf{B}_i)\|_2} \right) \right], \quad (3)$$

式中: \mathbf{B}_i 和 \mathbf{B}_k 为动态样本库中的第 i 个和第 k 个动态样本且 $i \neq k$; t_{ps} 为当前的动态样本数。 $i \neq k$ 将相似度计算限制在不同的样本之间; 中括号内的左项表示所有动态样本之间的相似度之和, 共计 $t_{\text{ps}}(t_{\text{ps}} - 1)$ 个相似度, 在该项中每一对样本实际进行了两次计算; 而右项是初始样本与所有动态样本的相似度之和, 为使其相似度具有和左项相同的权重, 其值需乘以 2, 共计 $2t_{\text{ps}}$ 个相似度; 最终得到 $t_{\text{ps}}(t_{\text{ps}} + 1)$ 个相似度之和, 然后除以总数得到平均相似度。该方法能够根据目标自身的运动特点产生阈值, 若目标在跟踪过程中变化得比较剧烈, 则该阈值较小; 若目标变化得越平缓, 则该阈值越大。成功帧的目标样本将被收录到动态样本库中, 当样本总数超过 t_z 时, 最早的动态样本将被删除, 而最具准确性和完整性的初始样本被永久保留。

2.3 多任务学习策略

为使跟踪器建立完整的目标外观模型, 本文采

用多任务学习方式使网络并行学习多个相关联的子任务^[18], 提高训练效率。跟踪器的第一个子任务是适应目标当前的外观。它需要跟踪器从最近几帧学习大量的样本来适应目标外观的最新变化, 更新模型对目标的短期记忆。短期记忆损失是二分类交叉熵损失函数, 表示为

$$L_{\text{short}} = -y_p \log_e [g(x_p)] - y_n \log_e [g(x_n)], \quad (4)$$

式中: x_p, y_p 为跟踪过程中收集的短期正样本及其标签; x_n, y_n 为短期负样本和标签; g 为样本推理的输出值。

然而, 对短期样本的大量学习易使跟踪器过拟合于目标的当前外观, 导致其对目标之前外观的遗忘。因此, 跟踪器的第二个子任务是在长期维度上建立对目标外观的全面记忆。本文采用重演的方式, 即让跟踪器定期重复学习目标之前的外观^[19], 减少其对目标外观特征的遗忘。结合 2.2 节动态样本库中长期收集到的目标外观样本, 长期记忆损失

函数为

$$L_{\text{long}} = -y_1 \log_e [g(x_1)] - y_n \log_e [g(x_n)], \quad (5)$$

式中: x_1 、 y_1 为动态样本库中的长期正样本及其标签。

联合了对目标外观长期记忆与短期记忆的多任务损失函数为

$$L_{\text{mt}} = \lambda_1 L_{\text{short}} + \lambda_2 L_{\text{long}}, \quad (6)$$

式中: λ_1 是短期记忆损失权重; λ_2 是长期记忆损失权重。多任务学习通过平衡短期和长期记忆缓解了跟踪器对目标外观的遗忘,使其在适应目标最新变化的同时能够在长期维度上建立完整的目标外观模型。

2.4 快速重检测

重检测分为区域提议和分类评估两步。为加快重检测速度,本文根据物体的轮廓特征和尺度信息进行高质量的区域提议。跟踪目标往往具有较高的实例属性,即具有很多轮廓线段,而纯粹背景中的轮廓较稀疏。因此本文引入 Edge Boxes 方法^[20]来产生大量具有较高实例属性的边界框,在一帧中产生 N_{gl} 个提议区域。并且跟踪场景中往往存在多个实例对象,为减少其对后续分类评估带来的负担,本文

根据目标的尺度信息排除一部分提议区域。成功的跟踪结果的尺度信息会被收集。在重检测过程中提议的区域,需满足以下条件,

$$\begin{cases} \omega_o - 2(\omega_o - \omega_{\min}) < \omega_j < \omega_o + 2(\omega_{\max} - \omega_o) \\ h_o - 2(h_o - h_{\min}) < h_j < h_o + 2(h_{\max} - h_o) \end{cases}, \quad (7)$$

式中: ω_{\max} 和 ω_{\min} 分别是所有成功跟踪结果的最大宽度和最小宽度; ω_o 是初始帧目标框的宽度; h_{\max} 和 h_{\min} 分别是所有成功跟踪结果的最大高度和最小高度; h_o 是初始帧目标框的高度; ω_j 和 h_j 分别是当前全图区域提议中第 j 个目标框的宽和高。(7)式以最可信的初始尺度为基础,同时考虑了尺度变化以及区域提议时包围框的误差。

2.5 跟踪方法的实施

本文跟踪方法的实施流程如表 1 所示。在跟踪序列开始的前 h 帧,由于动态样本库中的样本太少,仅由置信度 c_{on} 检测跟踪结果。为了减少跟踪器在目标剧烈变化时漂移至其他物体上的可能,在连续丢失目标 r 帧后再进行全图搜索。同时每 b 帧进行一次多任务学习,保持跟踪器对目标外观的全面记忆。

表 1 跟踪方法的实施流程

Table 1 Implementation flow of the tracking method

Algorithm 1: Proposed tracking method

Input: Initial target position v_1 , pretrained local tracker.

Output: Estimated target position v_q

- 1: Initial training the local tracker by the first frame, set number of consecutive failures $P_{\text{cf}}=0$.
- 2: **for** $q=2:m$
- 3: **if** $P_{\text{cf}} \leq r$ **then** locally draw candidate regions of target
- 4: **else** globally draw candidate regions of target
- 5: Get the tracking result and confidence score using local tracker
- 6: **if** ($c_{\text{on}} > 0$ **and** $s_{\text{cos}} > H_{\text{cos}}$ **and** $q > h$) **or** ($c_{\text{on}} > 0$ **and** $q \leq h$) **then**
 $P_{\text{cf}}=0$, collect new short-term and long-term samples
- 7: **else** $P_{\text{cf}}=P_{\text{cf}}+1$
- 8: **if** ($q \% b \neq 0$) **then** update local tracker using Eq. (4)
- 9: **if** ($q \% b = 0$) **then** update local tracker using Eq. (6)

3 实验与分析

3.1 实验平台、参数与数据集

实验的软件平台为 Windows 10 操作系统, pytorch 1.0 版本。硬件平台为 Intel 9400F

2.90 GHz CPU, 具有 8 GB 显存的 RTX2080 GPU 以及 16 GB 的 RAM。本文的跟踪方法在该平台上以 3 frame/s 的速度运行。

本文在 ImageNet-Vid 数据集^[21]上对网络进行预训练,并参考文献[5]设置在线跟踪过程中网络模

型的超参数,使本地跟踪器获得良好的特征提取与分类能力。损失权重 λ_1 和 λ_2 均为 1。全图提议的数量 $N_{gl} = 256$ 。帧数阈值中 $r = 3, h = 16, b = 10$ 。动态样本库的容量 t_z 由实验确定。

本文使用 OTB-2015 数据集^[22] 和 VOT-2016 数据集^[23] 测试跟踪方法的效果。OTB-2015 含有 100 个视频序列,分为 11 种跟踪场景。该数据集通过计算一次通过评估(OPE)的精确度和成功率评价跟踪器的性能。精确度是中心位置误差小于 20 pixel 的帧数的百分比。成功率是交并比小于不同阈值的曲线的积分面积。VOT-2016 数据集由 60 个视频组成,评价指标是精度、鲁棒性和平均重叠(AO)分数。在精度和鲁棒性测试中,每次跟踪失败都会重新初始化跟踪器;而 AO 分数则是 OPE 测试中的平均交并比。

3.2 目标跟踪实验

动态样本库的容量 t_z 影响着跟踪异常检测的准确性和多任务学习的质量。图 2 展示了动态样本库容量对跟踪效果的影响。在容量从 16 增加到

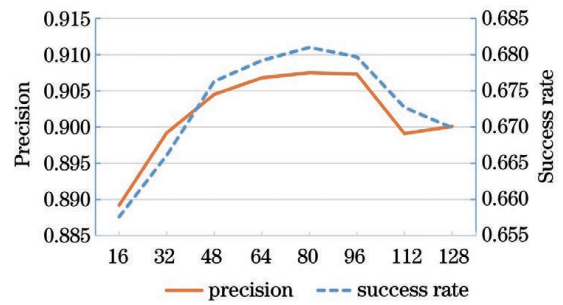


图 2 不同 t_z 值下的跟踪结果

Fig. 2 Tracking results under different t_z values

128 的过程中,该方法的效果呈现先提高后降低的趋势,在容量为 80 时跟踪效果达到了最佳。本文将 t_z 设置为 80。

为了研究各个改进方案对跟踪性能的影响,本文在 OTB-2015 数据集上进行了消融研究,结果如表 2 所示。在仅加入丢失检测步骤时,由于缺少高质量的重检测方法,在成功率提高的同时精度反而略微下降。而在此基础上引入其他两项改进后,精度和成功率得到明显提高。

表 2 消融实验结果

Table 2 Results of ablation experiments

Tracker	Detection of loss	Multi-task learning	Fast re-detection	Precision	Success rate
Baseline				0.9016	0.6677
Method 1	✓			0.8986	0.6722
Method 2	✓	✓		0.9061	0.6788
Method 3	✓		✓	0.9032	0.6758
Method 4	✓	✓	✓	0.9075	0.6812

将本文所提跟踪器与一些经典的跟踪器进行比较,结果如图 3 所示。其中空间正则化相关滤波器(SRDCF)^[11]、模板与像素互补学习方法(Staple)^[8] 是基于相关滤波类的判别类跟踪器,多域卷积神经

网络(MDNet)^[5] 是基于深度学习的判别类跟踪器,孪生候选区域提议网络(SiamRPN)^[6]、全卷积孪生网络(SiamFC)^[4] 是孪生网络跟踪器。本文提出的跟踪器在该数据集上的精确度为 0.908,成功

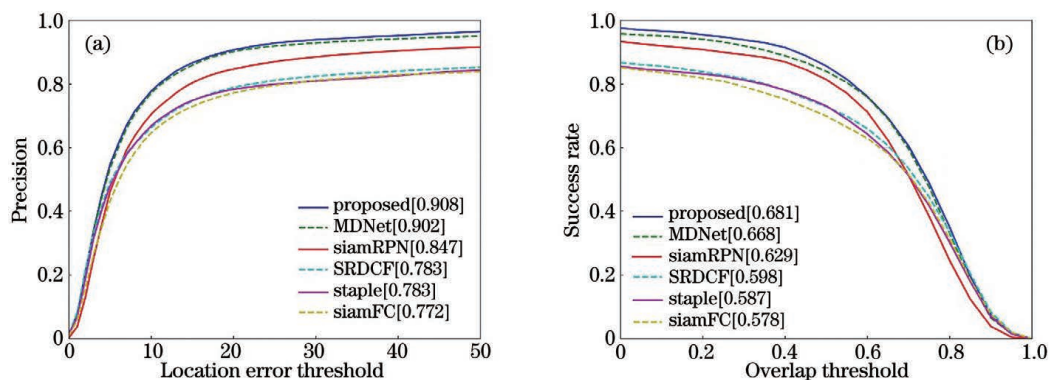


图 3 OTB-2015 数据集测试结果。(a)精度;(b)成功率

Fig. 3 Testing results of the OTB-2015 dataset. (a) Precision; (b) success rate

率为 0.681, 在 6 种跟踪方法中具有最佳的综合表现。

表 3 展示了这些跟踪器在 11 种不同挑战属性中的精确度。本文方法在目标被部分遮挡(OCC)和背景杂乱(BC)的场景中具有领先的性能, 并且在目标发生尺度变化(SV)、形变(DEF)、平面内旋转

(IPR)、平面外旋转(OPR)时具有明显的优势。而在环境光照变化(IV)、目标快速运动(FM)以及目标超出视野(OV)的挑战中, 本文方法也具有较好的成绩。MDNet 在运动模糊(MB)挑战中具有更佳的性能。SiamRPN 善于在低分辨率(LR)的视频序列中进行跟踪。

表 3 11 种挑战属性中的跟踪精度

Table 3 Tracking precision of 11 challengeable attributes

Tracker	IV	SV	OCC	DEF	MB	FM	IPR	OPR	OV	BC	LR
Proposed	0.909	0.889	0.858	0.876	0.854	0.877	0.907	0.898	0.838	0.937	0.937
MDNet	0.907	0.874	0.828	0.866	0.858	0.872	0.898	0.883	0.833	0.907	0.945
SiamRPN	0.859	0.838	0.780	0.826	0.816	0.789	0.854	0.851	0.725	0.799	0.978
SRDCF	0.792	0.745	0.734	0.735	0.765	0.768	0.745	0.741	0.593	0.775	0.760
SiamFC	0.736	0.736	0.723	0.691	0.707	0.744	0.743	0.758	0.672	0.692	0.900
Staple	0.783	0.726	0.728	0.752	0.698	0.708	0.768	0.737	0.664	0.749	0.690

图 4 为 6 种跟踪方法在 4 组视频序列的跟踪结果。Bird1 序列的挑战因素主要是形变和超出视野。在跟踪过程中, 目标被云雾完全掩盖, 从而消失在视野内, 所有跟踪器丢失目标。而目标重新出现

后, 本文方法通过高质量的重检测环节, 精确地找回并跟踪目标。Matrix 序列的挑战因素主要是光照变化、快速运动和背景杂乱。在目标人脸快速运动的情况下, 背景中还含有大量相似的人脸, 所有跟踪



图 4 4 组视频序列的跟踪结果。(a) Bird1; (b) Box; (c) Girl2; (d) Matrix

Fig. 4 Tracking results of four video sequences. (a) Bird1; (b) Box; (c) Girl2; (d) Matrix

器都产生了漂移现象。而本文的跟踪器在发生漂移时减少了对嘈杂样本的学习,能重新找到正确的目标位置。Box 序列的挑战因素主要是尺度变化、遮挡和背景杂乱。当目标被部分遮挡时,其他跟踪器容易漂移到背景中的其他黑色物体上。而本文的跟踪器可以检测出异常的跟踪结果,在杂乱的背景中依然有良好的性能。Girl2 序列的挑战因素主要是形变、运动模糊和遮挡。在 110 帧附近目标被完全

遮挡,使得跟踪器丢失目标。而本文的跟踪器通过丢失检测避免了遮挡情况下对目标样本的学习,可以在目标重新出现后快速找到目标。

本文还在 VOT-2016 数据集上测试了跟踪效果,结果如表 4 所示。连续卷积算子(C-COT)^[24]方法具有最佳的鲁棒性。而本文算法在更接近实际跟踪情况的 OPE 测试中具有明显的优势,同时具有较好的精度和鲁棒性。

表 4 VOT-2016 数据集测试结果

Table 4 Testing result of the VOT-2016 dataset

Tracker	C-COT	Staple	MDNet	SiamFC	SRDCF	Proposed
Accuracy	0.539	0.544	0.541	0.532	0.535	0.545
Robustness	0.238	0.378	0.337	0.461	0.419	0.317
AO	0.469	0.388	0.457	0.399	0.397	0.482

4 结 论

针对跟踪器受到嘈杂样本影响的问题,本文对跟踪过程中的异常检测和在线更新步骤做出改进。该方法通过比较样本间的相似度减少了对嘈杂样本的获取,并使用动态阈值来适应变化剧烈程度不同的跟踪目标,改善了跟踪器漂移的问题;采用多个子任务的加权损失函数来联合长期、短期样本的训练,保持了跟踪器对目标外观长期而全面的记忆,提高了复杂场景下的跟踪稳定性;在全图搜索时根据实例属性进行区域提议,并根据尺度信息减少无效提议,提高了找回目标的效率。与各种经典的跟踪方法相比,该跟踪方法对长期跟踪过程中出现的遮挡、背景杂乱等挑战均具有稳健的跟踪效果。

参 考 文 献

- [1] Sunkara J K, Santhosh M, Cherukuri S B, et al. Object tracking techniques and performance measures: a conceptual survey [C] // 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), September 21-22, 2017, Chennai, India. New York: IEEE Press, 2017: 2297-2305.
- [2] Li P X, Wang D, Wang L J, et al. Deep visual tracking: review and experimental comparison [J]. Pattern Recognition, 2018, 76: 323-338.
- [3] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 2544-2550.
- [4] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking [M] // Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [5] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4293-4302.
- [6] Li B, Yan J J, Wu W, et al. High performance visual tracking with Siamese region proposal network [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8971-8980.
- [7] Sun Y J, Zhang L Y, Yun X. Visual tracking algorithm based on region estimation and adaptive classification [J]. Laser & Optoelectronics Progress, 2019, 56(18): 181001.
孙彦景, 张丽颖, 云霄. 基于区域预估与自适应分类的视觉跟踪算法 [J]. 激光与光电子学进展, 2019, 56(18): 181001.
- [8] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1401-1409.
- [9] Danelljan M, Häger G, Khan F S, et al. Convolutional features for correlation filter based

- visual tracking[C]//2015 IEEE International Conference on Computer Vision Workshop (ICCVW), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 621-629.
- [10] Liu H F, Sun C, Liang X L. Correlation-filter tracking algorithm with adaptive-feature fusion and anti-occlusion[J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141014.
刘海峰, 孙成, 梁星亮. 自适应特征融合与抗遮挡的相关滤波跟踪算法[J]. *激光与光电子学进展*, 2020, 57(14): 141014.
- [11] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4310-4318.
- [12] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.
- [13] Yan B, Zhao H J, Wang D, et al. 'Skimming-perusal' tracking: a framework for real-time and robust long-term tracking [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 2385-2393.
- [14] Wang Z P, Wang H, Fang B F, et al. Support vector correlation filter with long-term tracking[J]. *Signal, Image and Video Processing*, 2018, 12(8): 1541-1549.
- [15] Shen Y L, Wu Z D, Zhao R J, et al. Long-term object tracking based on model updating and fastre-detection[J]. *Acta Optica Sinica*, 2020, 40(3): 0315002.
沈玉玲, 伍忠东, 赵汝进, 等. 基于模型更新与快速重检测的长时目标跟踪[J]. *光学学报*, 2020, 40(3): 0315002.
- [16] Matthews L, Ishikawa T, Baker S. The template update problem[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(6): 810-815.
- [17] Parisi G I, Kemker R, Part J L, et al. Continual lifelong learning with neural networks: a review[J]. *Neural Networks*, 2019, 113: 54-71.
- [18] Cipolla R, Gal Y, Kendall A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7482-7491.
- [19] Gepperth A, Karaoguz C. A bio-inspired incremental learning architecture for applied perceptual problems [J]. *Cognitive Computation*, 2016, 8(5): 924-934.
- [20] Zitnick C L, Dollár P. Edge boxes: locating object proposals from edges[M]//Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8693: 391-405.
- [21] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [22] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834-1848.
- [23] Kristan M, Leonardis A, Matas J, et al. The visual object tracking VOT2016 challenge results[M]//Hua G, Jégou H. *Computer vision-ECCV 2016 workshops. Lecture notes in computer science*. Cham: Springer, 2016, 9914: 777-823.
- [24] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9909: 472-488.